

Human Action Recognition Based on Key Postures

A Thesis Submitted for the Degree of
Doctor of Philosophy

By
Yan Chen

in

School of Computing and Communications
UNIVERSITY OF TECHNOLOGY, SYDNEY
AUSTRALIA
SEPTEMBER 2009

UNIVERSITY OF TECHNOLOGY, SYDNEY
SCHOOL OF COMPUTING AND COMMUNICATIONS

The undersigned hereby certify that they have read this thesis entitled "**Human Action Recognition Based on Key Postures**" by **Yan Chen** and that in their opinions it is fully adequate, in scope and in quality, as a thesis for the degree of **Doctor of Philosophy**.

Dated: September 2009

Research Supervisor:



Xiangjian He

CERTIFICATE

Date: September 2009

Author: **Yan Chen**

Title: **Human Action Recognition Based on Key Postures**

Degree: **Ph.D.**

I certify that this thesis has not already been submitted for any degree and is not being submitted as part of candidature for any other degree.

I also certify that the thesis has been written by me and that any help that I have received in preparing this thesis, and all sources used, have been acknowledged in this thesis.



Signature of Author

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to my supervisor, A/Prof. Xiangjian He for his continuous encouragement, advice, help and invaluable suggestions. I feel really happy, comfortable and unconstrained with him during my PhD study here. Many thanks are also due to my co-supervisors, A/Prof. Tom Hintz and Dr. Qiang Wu for their valued suggestions and constant supports, and for the numerous conversations with them. I gratefully acknowledge the invaluable discussion with Prof. Massimo Piccardi.

I appreciate the financial assistance of Australian Postgraduate Award (APA) provided by the Australian government. I appreciate the travel support for attending the conferences which I received from the previous Faculty of Information Technology and the UTS Vice-Chancellor's Conference Fund. I also appreciate the financial support from a Student Equity Grant of UTS, which helped me release from housework.

I wish to thank my fellow research students and the staff of the school, especially those people listed below for providing various assistance for the completion of this research work.

- Wenjing Jia, Lihong Zheng, Huaifeng Zhang, Oscar Perez Concha, Hatice Gunes, Ahmed Otoom, Zui Zhang, Ruo Du, Cao Zeng .

Last but not least, I would like to thank my husband, Jianming, for his understanding and assistance. I also thank my parents and parents-in-law for their constant support. This thesis could not have been completed without the support and encouragement of my brother and my best friends.

To My Family, Felix and Ivan

Table of Contents

| | |
|---|----------|
| Table of Contents | vi |
| List of Tables | ix |
| List of Figures | x |
| Author's Publications for the Ph.D | xiii |
| Abstract | xv |
| 1 Introduction | 1 |
| 1.1 Application Domains of Human Motion Analysis | 2 |
| 1.2 Research Motivation | 3 |
| 1.3 Author's Contributions | 4 |
| 1.4 Thesis Organization | 5 |
| 1.5 Terminologies | 7 |
| 2 Literature Review of Human Motion Analysis | 8 |
| 2.1 Human Detection | 10 |
| 2.1.1 Motion Segmentation | 10 |
| 2.1.2 Moving Object Classification | 15 |
| 2.2 Human Tracking | 16 |
| 2.3 Human Behaviour Understanding | 19 |
| 2.3.1 Human Posture Representation | 20 |
| 2.3.2 Human Action Representation | 23 |
| 2.3.3 Action Recognition | 27 |
| 2.3.4 Knowledge of Key Posture Based Action Recognition | 31 |
| 2.4 Summary | 34 |

| | | |
|----------|--|------------|
| 3 | Human Detection | 36 |
| 3.1 | Moving Object Extraction | 37 |
| 3.1.1 | Adaptive MoG Model | 38 |
| 3.2 | Pedestrian Detection in Video | 40 |
| 3.2.1 | Pedestrian Recognition Using Histogram of Angle | 41 |
| 3.2.2 | Experiments | 46 |
| 3.3 | Human Detection in Static Images | 51 |
| 3.3.1 | Hausdorff Distance | 51 |
| 3.3.2 | Mahalanobis Distance | 53 |
| 3.3.3 | Mahalanobis Distance VS Hausdorff Distance | 54 |
| 3.3.4 | Human Detection Based on Human Geometrical Structure | 54 |
| 3.3.5 | Experiments | 63 |
| 3.4 | Summary | 68 |
| 4 | Human Action Recognition Based on Shape Features | 70 |
| 4.1 | Key Posture Selection | 71 |
| 4.1.1 | Coarse Selection | 72 |
| 4.1.2 | Fine Selection | 76 |
| 4.2 | Action Recognition | 83 |
| 4.2.1 | Dynamic Time Warping | 84 |
| 4.2.2 | Action Template Creation and Matching | 89 |
| 4.3 | Experiments | 98 |
| 4.4 | Summary | 99 |
| 5 | Human Action Recognition Based on Radon Transform | 102 |
| 5.1 | Human Posture Representation By Radon Transform | 103 |
| 5.1.1 | Radon Transform | 103 |
| 5.1.2 | Human Posture Representation Using Radon Transforms | 105 |
| 5.2 | Key Posture Selection | 106 |
| 5.2.1 | Affinity Propagation Clustering | 106 |
| 5.2.2 | Key Posture Identification | 110 |
| 5.3 | Action Recognition | 110 |
| 5.3.1 | Action Template Creation | 111 |
| 5.3.2 | Learning and Classification Procedure | 113 |
| 5.4 | Experiments | 117 |
| 5.4.1 | Experiments 1 | 117 |
| 5.4.2 | Experiments 2 | 119 |
| 5.5 | Summary | 126 |

6 Conclusions and Future Work 129

6.1 Conclusions 129

6.1.1 Human Detection 129

6.1.2 Human Action Recognition 131

6.2 Future Work 133

Appendix A

Pictures Used in Chapter 3 135

Appendix B

Part of Action Videos 142

Bibliography 152

List of Tables

| | | |
|------|---|-----|
| 3.1 | Experiments results for pedestrian detection | 50 |
| 3.2 | Detail accuracy for pedestrian detection | 50 |
| 4.1 | Confusion matrix for action recognition | 100 |
| 4.2 | Accuracies for action recognition | 100 |
| 5.1 | Confusion matrix for leave-one-subject-out cross validation (SMO) . . | 119 |
| 5.2 | Accuracies for leave-one-subject-out cross validation (SMO) | 120 |
| 5.3 | Confusion matrix for leave-one-subject-out cross validation (BayesNet) | 120 |
| 5.4 | Accuracies for leave-one-subject-out cross validation (BayesNet) . . . | 121 |
| 5.5 | Confusion matrix for leave-one-subject-out cross validation (C4.5) . . | 121 |
| 5.6 | Accuracies for leave-one-subject-out cross validation (C4.5) | 122 |
| 5.7 | Confusion matrix for leave-one-out cross validation (SMO) | 123 |
| 5.8 | Detail accuracy for leave-one-out cross validation(SMO) | 123 |
| 5.9 | Confusion matrix for leave-one-out cross validation (BayesNet) | 124 |
| 5.10 | Detail accuracy for leave-one-out cross validation (BayesNet) | 124 |
| 5.11 | Confusion matrix for leave-one-out cross validation (C4.5) | 125 |
| 5.12 | Detail accuracy for leave-one-out cross validation (C4.5) | 125 |
| 5.13 | Comparison with related studies | 126 |

List of Figures

| | | |
|------|--|----|
| 1.1 | Framework of human action recognition based on key postures | 6 |
| 2.1 | Framework human motion analysis [109] | 9 |
| 2.2 | General model of articulated human body kinematic [26] | 21 |
| 2.3 | Human posture representation of body part [7] | 24 |
| 2.4 | The posture representation [59] | 26 |
| 2.5 | An example of Hidden Markov Models | 28 |
| 3.1 | Cycle of human walking | 42 |
| 3.2 | Outline of pedestrian detection | 42 |
| 3.3 | Object and its skeleton | 43 |
| 3.4 | Angle histograms between centroid point and bottom end points before normalization. (a) Pedestrian angle histograms, (b) Car angle histograms | 48 |
| 3.5 | Angle histograms between centroid and two end bottom points after normalization: (a) Pedestrian angle histograms. (b) Car angle histograms | 49 |
| 3.6 | Variance of angle histogram | 50 |
| 3.7 | The Hausdorff distance calculation example | 53 |
| 3.8 | Separating image into blocks [105] | 55 |
| 3.9 | Example of distance map. (a): Distance map of human images; (b): Distance map of non-human images | 57 |
| 3.10 | Average distance map. (a): human; (b): non-human | 58 |
| 3.11 | Difference map between human and non-human average map | 59 |
| 3.12 | Data flow for calculation projection matrix | 61 |

| | | |
|------|---|----|
| 3.13 | Image examples used in the experiments. (a) non-human images; (b) human images | 63 |
| 3.14 | Score distribution obtained by different methods: a) human images with Mahalanobis distance; b) non-human images with Mahalanobis distance; c) human images with Hausdorff distance; d) non-human images with Hausdorff distance. | 65 |
| 3.15 | Recognition rates obtained under different parameters with different methods | 66 |
| 3.16 | Ratios of inter-class scatter and intra-class scatter obtained under different parameters with different methods | 67 |
| 4.1 | Block devision of a human image | 73 |
| 4.2 | Information measurement of a running sequences | 74 |
| 4.3 | Coarse selection work flow. | 75 |
| 4.4 | Key posture candidates after coarse selection (running): (a) Original sequence; (b) T=10; (c) T=5 | 77 |
| 4.5 | Key posture candidates after coarse selection (Jack): (a) Original sequence; (b) T=10; (c) T=5 | 78 |
| 4.6 | Eight-point representation of a human body: (a) Interested points selecting; (b) Points labeling | 79 |
| 4.7 | Final key postures of running action: (a) T=10; (b) T=5 | 83 |
| 4.8 | Final key postures of jack action: (a) T=10; (b) T=5 | 84 |
| 4.9 | Examples of final key postures for actions in database | 85 |
| 4.10 | A warping between two sequences | 86 |
| 4.11 | Warping path example | 87 |
| 4.12 | An example of distance matrix | 90 |
| 4.13 | An example of a warping path cost matrix and its warping path . . . | 90 |
| 4.14 | Distance calculation between two sequences | 92 |
| 4.15 | Distance matrix between sequence Q and sequence R | 94 |
| 4.16 | Warping path calculation example—step 3 | 94 |

| | | |
|------|---|-----|
| 4.17 | Warping path calculation example—step 4 | 95 |
| 4.18 | Warping path calculation example—step 5 | 95 |
| 4.19 | Warping path calculation example—step 6(1) | 96 |
| 4.20 | Warping path calculation example—step 6(2) | 96 |
| 4.21 | Warping path calculation example—step 6(3) | 97 |
| 4.22 | Warping path for the example | 97 |
| 4.23 | An example of database(jack) | 99 |
| | | |
| 5.1 | The Radon transform computation [1] | 103 |
| 5.2 | Radon transforms for two parallel lines | 104 |
| 5.3 | Human images and their corresponding Radon transforms | 106 |
| 5.4 | Affinity propagation clustering procedure | 109 |
| 5.5 | A linear Support Vector Machine [83] | 115 |
| 5.6 | An example of Bayesian network | 116 |
| 5.7 | An example of Decision Tree | 117 |

Author's Publications for the Ph.D

1. **Y. Chen**, Q. Wu, and X. He, "Human Action Recognition by Radon Transform," in book 'Multimedia Analysis, Processing and Communications', 2009, accepted.
2. **Y. Chen**, Q. Wu, and X. He, "Human Action Recognition by Radon Transform," in 2008 IEEE International Conference on Data Mining Workshops (ICDMW2008), pp.862-868, 2008.
3. **Y. Chen**, Q. Wu, and X. He, "Extracting Key Postures Using Radon Transform," in Image and Vision Computing New Zealand (IVCNZ08), IEEE Digital Library, 5 pages, 2008.
4. C. Du, Qiang.Wu., J. Yang, X. He, **Y. Chen**, "Subspace Analysis Methods Plus Motion History Image for Human Action Recognition", in Digital Image Computing: Techniques and Applications (DICTA2008), Canberra, Australia, pp. 606-611, December 2008.
5. **Y. Chen**, Q. Wu, X. He, W. Jia, and T. Hintz, "A Modified Mahalanobis Distance for Human Detection in Out-door Environments," in the First IEEE International Conference on Ubi-Media Computing, 2008, pp. 243-248.
6. **Y. Chen**, Q. Wu, X. He, C. Du, and J. Yang, "Extracting Key Postures in a Human Action Video Sequence," in the 10th International Workshop on Multimedia Signal Processing (MMSP2008), 2008, pp. 569-573.

7. **Y. Chen**, Q. Wu, and X. He, "Using Dynamic Programming to Match Human Behavior Sequences," in the 10th International Conference on Control, Automation, Robotics and Vision (ICARV2008) pp.1498-1503, 2008.
8. **Y. Chen**, Q. Wu, and X. He, "Motion Based Pedestrian Recognition," in IEEE Congress on Image and Signal Processing, 2008, pp. 376-380.
9. Y. Jiang, **Y. Chen**, Z. Zeng, and X. He, "A Bank Customer Credit Evaluation Based on the Decision Tree and the Simulated Annealing Algorithm," in the 8th IEEE International Conference on Computer and Information Technology, 2008, pp. 203-206.
10. X. He, J. Li, and **Y. Chen**, "Local Binary Patterns with Mahalanobis Distance Maps for Human Detection," in IEEE Congress on Image and Signal Processing, 2008, pp. 520-524.
11. **Y. Chen**, Q. Wu, X. He, W. Jia, and T. Hintz, "Pixel Structure Based on Hausdorff Distance for Human Detection in Out-door Environments," in Digital Image Computing: Techniques and Application (DICTA2007), pp. 67-72, 2007.
12. **Y. Chen**, Q. Wu, X. He, W. Jia, and T. Hintz, "Study on Human Behaviour Retrieval," in Proceeding of International Conference on Image Processing, Computer Vision, and Pattern Recognition, pp. 448-454, 2007.

Abstract

Human motion analysis has gained considerable interests in the computer vision area due to the large number of potential applications and its inherent complexity. Currently, human motion analysis is at an early stage. Its final aim is to generate an easy understanding, high level semantic description in a given scene. Human action recognition is an important step to the final aim of human motion analysis.

Human Detection

Human detection is part of the field of human motion analysis. The thesis looks at human detection. The thesis proposes a method using histogram of angles to discriminate pedestrians from vehicles. This proposed method is encouraged by the reality that humans are non-rigid objects. An angle formed by the centroid point and two bottom points for a human changes periodically while the angle for the vehicle is relatively static. In this part, this thesis also presents an approach to detect humans in static images. The thesis proposes an approach which uses human geometric features to fulfill the task.

Human Action Recognition

The thesis focuses on human action recognition. The thesis proposes what will be called a key postures based human action recognition approach. As we have known, human actions can be well described by a few important postures (called key postures) which are significantly different from each other and all other postures can be

clustered to these key postures. Therefore, these key postures can be used to represent and to infer the corresponding human action. The benefit of using key postures to represent human action is to reduce computational complexity. The thesis proposes two methods for human action recognition based on key postures. One is a human action recognition based on shape features and the other one is action recognition based on Radon transforms. Both methods follow three steps to achieve action recognition. These steps are video processing, key posture extraction and action recognition.

A two-step approach is proposed to extract key postures from preprocessed action video. These two steps are coarse selection and fine selection. Feature extraction and representation are discussed in both steps. After key postures are extracted from a video, key posture sequences are used to represent human actions. Each key posture sequence is regarded as an action template. In order to compare two action sequences, Dynamic Time Warping (DTW) is applied to determine the distance between the two action sequences.

In the second method, in order to obtain key postures, the action sequences are extracted from the preprocessed silhouettes using Radon transforms. Then, an unsupervised cluster analysis is applied to Radon transforms to identify the key postures for each sequence. Such key postures are used in the subsequent training and testing procedure. Several benchmark classifiers are used in this work for action learning and classification.

Author's Publications

This thesis covers the research results conducted by the author while undertaking for the degree. Most of the results have been published in research papers in refereed publications which are listed in Author's Publication for Doctor of Philosophy (PhD).